AD A056967

AD No.
DDC FILE COPY

11

TECHNOLOGY FOR DIRECTOR DUBIOUS: EVALUATION AND DECISION IN PUBLIC CONTEXTS

SOCIAL SCIENCE RESEARCH INSTITUTE
UNIVERSITY OF SOUTHERN CALIFORNIA

Ward Edwards

LEVEL II

This document has been approved for public release and sale; its distribution is unlimited.

ADVANCED ARPA DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

The objective of the Advanced Decision Technology Program is to develop and transfer to users in the Department of Defense advanced management technologies for decision making. These technologies are based upon research in the areas of decision analysis, the behavioral sciences and interactive computer graphics. The program is sponsored by the Cybernetics Technology Office of the Defense Advanced Research Projects Agency and technical progress is monitored by the Office of Naval Research — Engineering Psychology Programs. Participants in the program are:

Decisions and Designs, Incorporated
Harvard University
Perceptronics, Incorporated
Stanford Research Institute
Stanford University
The University of Southern California

Inquiries and comments with regard to this report should be addressed to:

Dr. Martin A. Tolcott
Director, Engineering Psychology Programs
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

or

LT COL Roy M. Gulick, USMC
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

11

TECHNICAL REPORT SSRI 77-2

6

TECHNOLOGY FOR DIRECTOR DUBIOUS: EVALUATION AND DECISION IN PUBLIC CONTEXTS

by

10 Ward Edwards

9 Final technical rept. Oct 76-30 Sep 77

Sponsored by

Defense Advanced Research Projects Agency
ARPA Order No. 3052
Under Subcontract from
Decisions and Designs, Incorporated

12 44p.

11 August 1977

15 N00014-76-C-0074, ARPA Order-3052

SOCIAL SCIENCE RESEARCH INSTITUTE
University of Southern California
Los Angeles, California 90007

CONTENTS

Summary

This report examines the social, structural and organizational obstacles to the introduction of decision technology in public contexts, and summarizes two studies that suggest ways of overcoming these obstacles.

As a means of defining the problem the report caricatures two Federal government policy-makers: Director Devious and Director Dubious. Director Devious wants to keep his values and probabilities covert in order to enhance his freedom of action. Director Dubious, though a skeptic about new technologies, recognizes the problem they address and is willing to give them a try. Two classes of technological tools are proposed to him.

One technology, concerned with probability estimation and Bayesian inference, is illustrated by a study conducted by the American College of Radiology, using ARPA-developed technology, of the diagnostic value of x-rays. Attending physicians, minimally trained about probabilities, made pre- and post-x-ray probability judgments about possible diagnoses in emergency room cases. The log likelihood ratio inferred from these judgments was the measure of diagnosticity. The main conclusions were: (1) minimally trained physicians make very well calibrated probability estimates, (2) very few x-rays are completely undiagnostic, even if taken for medical-legal reasons, (3) level of physician training made little difference to performance in probability estimates.

The other technology, concerned with measurement of social values, is illustrated by an application of a version of multiattribute utility measurement to selection of nuclear waste disposal sites. Experts on nuclear waste disposal sites evaluated various hypothetical sites by an ARPA-developed procedure. The main findings were that they

liked the procedure and wanted to try it further, and that the results were robust under manipulations having to do with incorrect prior expectations concerning the ranges of dimensions of value.

Both technologies are offered to Director Dubious, and his governmental colleagues, as serious candidates for adaptation and use.

Figures

Tables

Acknowledgment

Preparation of this report and conceptual work on the place of decision analysis in public contexts was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract N00014-76-C-0074 under subcontract from Decisions and Designs, Inc. Data collection and analysis for empirical studies reported here were supported by Contract HSM 110-72-293 and Grant Number HS 01546 from the National Center for Health Services Research HRS, PHS/DHEW to the American College of Radiology and by the Joint IAEA-IIASA Project on Technological Risk Assessment. All empirical results reported here have been more extensively reported elsewhere, in other formats.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

Introduction

In preparing this paper, I had the enormous advantage of having read a related paper prepared by Mr. Joseph F. Coates, of the Office of Technology Assessment, U.S. Congress, (to appear in AAAS Symposium Volume on Judgment and Choice in Public Policy Decisions, in press). Mr. Coates's incisive and provocative analysis of the nature of public policy decision making and the difficulties that experts have in providing useful inputs to that process merits extravagant admiration. It is a frank, penetrating review of virtually all of the issues that bemused academics like myself who have fluttered around the fringes of the Federal policy community for many years have vaguely sensed as being characteristic of policy making.

I would like to underline a few points made by Mr. Coates, as a preliminary to some suggestions about what might be done to address them. Perhaps his most important single point is that policy is not made in a problem-oriented vacuum. Instead, it is made in an embattled arena, usually by a man or organization upon whom are focused the efforts of a wide variety of conflicting stake holders, each having his own perception of both problems and issues--often with his own collection of "facts" to back up that perception. As Mr. Coates says, "The key issue or issues are not obvious, since they usually have not been presented in a clear, cogent, or neutral way by any of the parties concerned. It is not in their interest to do so." In such an embattled context, "The resolution of an issue in almost all cases must be a compromise rather than a clear victory for any party to the conflict." This gladiatorial atmosphere presents problems to the would-be policy-influencer because "In general, experts cannot deal with trade-offs which are the essence of public policy. Experts cannot deal with compromise situations and conflict, as experts."

If one looks for the underlying issues of any conflict, they seem to fall into two categories: probabilities (measures of uncertainty) and utilities (measures of values) Concerning probabilities, Mr. Coates says "The future course of every public policy issue of necessity is involved in uncertainty. Much uncertainty is not accidental but intrinsic, and cannot be eliminated for several reasons. First, the future is not fully anticipatable; second, we do not have adequate models of social change; and third, many of the consequences of actions associated with policy cannot be understood until the actions themselves are taken." I would add that often those consequences cannot be understood even after the actions have been taken. As a result, Mr. Coates says that "Another primary task for government is to manage uncertainty, i.e., to take those measures that in one way or another eliminate, hedge, reduce, or compensate for uncertainty so as to permit the institutions of society to move ahead in an organized fashion." From my own point of view, such measures for uncertainty management have a necessary preliminary; first one must measure uncertainty.

The other issue that Mr. Coates identifies as crucial is the one that he calls value, but I would prefer for history-of-science reasons to call utility. He says, "The subject of values has engendered an alarming amount of intellectual trash, useless discussion, uninformed deliberation, and pointless hand wringing..... . Values are difficult to discern. Individuals often cannot see their own; when they can see them, they cannot give weights to them. Values are often ill-formed. They are latent, they are dark, they cannot necessarily be related to public decisions without a great deal of intermediate work."

On the question of measuring values, Mr. Coates seems to me to be somewhat ambivalent. At one point he says, "Since values are heterogeneous and overlapping among the parties of interest, it is difficult to identify and sort them into tidy bundles. An effective way to reveal the values of the parties to the conflict is important. That revelation is not likely to result from simple direct inquiry." At another point, he derides "...the false conclusion that making those values explicit is a worthwhile activity in all public policy processes..... . Many private motives are in conflict, are latent, are dark, uncongenial, and even unspeakable. Consequently the universal call for making them explicit in public is really an invitation to hypocrisy."

From reading Mr. Coates's paper, one can formulate a picture of two different Federal Government policy-makers, whom I shall call Director Devious and Director Dubious. Mr. Coates describes Director Devious quite well. "The crucial question facing public policy in any given time is striking a fresh balance among conflicting forces..... . The search for information is often a delaying tactic. It can be a mechanism for apparently taking action while taking no action..... . Even those most intimately associated with the issues.....often find it to their advantage not to confront (them), not to define them, not state them clearly, and not to use them as a basis for discourse, analysis, evaluation, and decision making..... . There is a tendency to misunderstand the role of the elected official and the senior decision maker in wanting him to make his values explicit. For him to make his values explicit would be a travesty. The decision makers role is to adjudicate and to keep his values internal so he can affectively adjudicate the value-laden material put forward to him by others."

I have much more difficulty in finding Mr. Coates's paper a description of Director Dubious. Mr. Coates says "Government is not a religion and bureaucrats are not moral athletes." But I believe that, in this as in other areas of performance, a desire for athletic excellence is built into many of us, whatever the level of our capabilities for fulfilling that desire. My image of Director Dubious is that he is perplexed by the multiplicity of the uncertainties and the value orientations with which he must cope. While he recognizes the necessity of functioning as a middle-man mediating among conflicting stake holders with conflicting values, in the face of technological and political realities that are often rather vaguely and uncertainly defined, he genuinely would like to perform this function as best he can, and would welcome tools that might help him to do so. Nor, I think, would he endorse Mr. Coates's advice that he should keep his own values deeply hidden from others, and perhaps even from himself. If some of his values are, as Mr. Coates says, dark, uncongenial, and even unspeakable, he wishes they weren't. He would like to have some way of inspecting values, both his own and those of others, and attempting to make some kind of moral sense out of them in their relation to the facts of the problem.

If I may lapse for a moment into psychoanalytic jargon, perhaps Director Devious might be taken as a representation of the ego of one kind of elected official or senior decision maker. If so, perhaps Director Dubious is a representation of the same person's superego.

I feel reasonably confident that Mr. Coates would regard the tools that I am going to propose for use as idealistic and naive, and therefore unlikely to be of much use to a public policy maker. Contexts exist in which I would agree with him. Nevertheless, each of the two major tools I plan to discuss is in fact in current use in significant public decision making contexts. Unfortunately, I will not present examples of the actual application of those tools to public decisions. For one thing, many of the details of those applications as they now are in progress are classified or otherwise confidential. For another thing, even if they were not, the character of each detailed application is typically so complicated that any attempt to present the basic ideas at appropriate length would inevitably fail. Consequently, I will talk about two relatively simple tools, both currently in use, in contexts in which they obviously bear on public policy, and could be used by public policy makers, but so far have not been.

Evaluating Radiological Efficacy by Bayesian Methods

My first tool is addressed to the first of the two key problems that Mr. Coates identified: the problem of uncertainty. The work that I will be reporting comes from the Efficacy Study of the American College of Radiology, and is a collaborative effort involving Lee Lusted, Russell Bell, Harry Roberts, David Wallace, and myself, among a good many others. The funds supporting it came from the National Center for Health Services Research of the U.S. Public Health Service. For a report on the results so far, see Lusted, Bell, Edwards, Roberts, and Wallace (in press).

The essential purpose of the Efficacy Study is to explore the usefulness of the very large number of X-rays and other radiologic diagnostic procedures being carried out in the United States. This particular report is based on 7,976 case studies in various emergency room settings. The study is ongoing; ultimately, it hopes to explore something on the order of 60,000 cases in a very wide variety of settings for radiological practice.

Back in 1971 the American College of Radiology set up a Committee on Efficacy. Among its motives were a finding by Bell and Loop (1971) that an X-ray examination of the skull following a trauma was quite unlikely to show skull fracture unless certain signs and symptoms were present, and that the probability was even lower that the radiographic findings would affect patient management or the final outcome. Bell and Loop estimated that society was paying $7,650.00 per skull fracture found in patients X-rayed under those conditions, and they questioned whether the benefits were worth the cost. More generally, the ACR's Board of Chancellors had been concerned because the demand for radiological services was, and is, growing faster than the supply, even though costs were also increasing. No rational basis existed at that time, or now, for setting priorities for available radiologic services. Customarily the radiologist performs the radiographic examination that the attending physician requests whether or not the request is appropriate. Although some data do exist suggesting what X-ray examinations are appropriate under what conditions, most radiologists know that on occasion a physician will request a radiological examination that appears unnecessary and the radiologist receiving the request is likely to fulfill it.

At its first meeting in 1971, the ACR Committee of Efficacy, chaired by Professor Lee Lusted of the University of Chicago, attempted to formulate the problem of what efficacy was and how it might be measured. Three different conceptions of efficacy were proposed, varying both in relevance to the long range problem and in measurability. The most relevant, but also hardest to measure, has come to be called Efficacy-3. Efficacy-3 is long run efficacy from the patient's point of view;that is, a diagnostic procedure is Efficacious-3 if the patient is, in the long run, better off as a result of that procedure and its consequences than he would have been had it not been performed. Obviously, knowledge of long run outcomes is difficult to obtain, and knowledge of hypothetical long run outcomes for sequences of diagnostic and therapeutic procedures other than the one actually carried out is even more difficult to obtain. Consequently, the committee next considered Efficacy-2. A diagnostic procedure is Efficacious-2 if and only if the course of subsequent therapeutic action taken by the attending physician is different as a result of performance of the procedure than it would have been otherwise.

Obviously Efficacy-2 is easier to measure than Efficacy-3, since it refers only to events in the immediate future. However, one must still discover what would have been done had constraints existed that did not in fact exist, and that too presents measurement difficulties. So, as a final fallback position, the Committee chose to study Efficacy-1. A procedure is Efficacious-1 if and only if the procedure influences the diagnostic thinking of the attending physician. This definition turns out to lead to relatively straightforward measurements. All one must do is to discover what the attending physician was thinking at the time he ordered the X-ray, what he thinks at the time he receives the results, and compare the two; if they are different, the procedure is Efficacious-1, and the size of the difference measures the amount of efficacy.

How does one measure what the attending physician is thinking? An appropriate procedure is to collect judgments of the probabilities of possible diagnoses prior to the X-ray, and another set of judgments posterior to it. Then, by using Bayes' Theorem, one can calculate the extent to which opinion has been changed as a result of the X-ray. Bayes' theorem is a trivially simple fact about probability, and can be represented for our current purposes by the following equation: LFO = LIO + LLR. In this equation, LIO stands for Log Initial Odds. LFO stands for Log Final Odds, and LLR stands for Log Likelihood Ratio. The logarithmic form of Bayes's theorem is used here in order to make the relationship additive, and in order to make the measure of diagnostic efficacy, LLR, symmetric around 0. The mathematical details by means of which this form of Bayes's theorem can be translated into other forms, and by means of which probability judgments can be related to this equation, can be found in many places, for example, Edwards, Lindman, and Phillips (1965). The more recent developemnt of this technology has been supported by this and other ARPA projects (see e.g. Eils, Seaver, and Edwards, 1977), and is in use in various military and international-relations contexts.

Obviously, at the time he orders an X-ray an attending physician may be considering many hypotheses about what is wrong with the patient. To reduce this large set to a more manageable set, the study defined two diagnoses. One of them was the most important diagnosis, the one that the attending physician would be most eager not to miss. In most cases that would be a fracture or some other medically unpleasant state of affairs. The other diagnosis was the diagnosis considered most likely; very often that was "normal".

A pretest of procedures for measuring Efficacy-1 is reported in Thornbury, Fryback, and Edwards (1975).

Figure 1 shows the front of a typical data collection form. This was filled out by the attending physican as a part of the process of ordering an X-ray. Figure 2 shows the back of that same form, which was filled out by the same physician when the result of the X-ray was returned to him. I must emphasize that the attending physicians in this study were not specially chosen for expertise in probability. The study was geographically very widely distributed; radiological settings in emergency rooms all over the country were used. Radiologists who were willing to cooperate in the study were brought from those settings to Chicago where they received roughly two days worth of training about the nature of the study and about some rather elementary rules

Patient Name ______________________ Patient I. D. ______________________

Date of Birth ______________ Sex ______________ Case Number ______________

AMERICAN COLLEGE OF RADIOLOGY - EFFICACY STUDY: SKULL - EMERGENCY

PART I (TO BE COMPLETED BY CLINICIAN BEFORE RADIOLOGIC PROCEDURE)
(See CLINICIAN'S HANDBOOK for guidance in completing this form.)

A. Clinical Data: For each entry check one box. (Y-Yes, N-No, ?-Equivocal, NA-Not Ascertained)

Y	N	?	ND	WAS REPORTED
				Recent Trauma
				Recent Pain or Headache
				Focal Weakness or Numbness
				Seizure or Unconsciousness
				Abnormal Mentation
				Deafness, Tinnitus, Vertigo
				Recent Visual Problems
				Defective Speech or Expression
				Recent Nausea or Vomiting

Other ______________________ (Specify)

Y	N	?	ND	WAS FOUND
				Physical Evidence of Injury
				Disrupted or Deformed Bone
				Focal Somatic Neural Defect
				Bruit or Altered Pulse
				Abnormal Mentation
				Discolored Eardrum or Otorrhea
				Eye Signs of Brain Problem
				Other Cranial Nerve Dysfunction
				Abnormal Tendon Reflex

Other ______________________ (Specify)

B. What is your patient's PROBLEM that causes you to request this examination? ______________________

C. 1) For the problem in B, state the most important prospective DIAGNOSIS which prompts this procedure. ______________________

2) What are your odds or probability estimate that the diagnosis in "C-1" will prove correct? ________

D. 1) For the problem in B, state the most likely prospective DIAGNOSIS ("normal" may be used) which prompts this procedure (only if different than the diagnosis in C) ______________________

2) What are your odds or probability that the diagnosis in "D-1" will prove correct? ________

E. What is the one major reason for this procedure? (Check one box only)

- ☐ Prove part normal
- ☐ Confirm diagnosis
- ☐ Investigate diffuse suspicions
- ☐ Confirm no change
- ☐ Show change in disease or healing
- ☐ Assess length, position, etc.
- ☐ Institutional policy
- ☐ Teaching or research
- ☐ Medical-legal

Other ______________

F. Are you presently aware of patient's medical insurance status?

Not Aware ☐ Believe patient is: Insured ☐ Not Insured ☐

Your Name ______________________ (Please Print) and/or ACR I. D. Number ______________ Date Filled Out ______________

RETURN TO RADIOLOGY AFTER COMPLETING PART II

NOT A PART OF MEDICAL RECORD

FIGURE 2

PART II TO BE COMPLETED BY CLINICIAN AS SOON AS RADIOLOGIC RESULTS ARE KNOWN

G. Knowing the X-ray findings, now estimate the odds or probability that the:

1) "most important" diagnosis stated in "C-1" of Part I is correct ____________

2) "most likely" diagnosis stated in "D-1", if any, of Part I is correct ____________

H. Enter below any NEW diagnoses based on radiological findings?

1) most important new diagnosis ____________ Code: (_ _ . _ _ _)

2) most likely new diagnosis (include normal) ____________ Code: (_ _ . _ _ _)

____

Your Name ____________ (Please Print) and/or ACR I. D. Number ____________ Date Filled Out ________

SIGNIFICANT RADIOLOGIC FINDINGS (To be filled out by radiologist or referring physician):

____

____

____

____

____

TO BE COMPLETED BY RADIOLOGY

RADIOLOGIC PROCEDURE CODE: _ _ _ _ _

RADIOLOGIC DIAGNOSES CODES Dx1 _ _ . _ _ _ _ _ Dx2 _ _ . _ _ _ _

Dx3 _ _ . _ _ _ _ _

SETTING (check one)

☐ Screening ☐ Inpatient

☐ Emergency ☐ Outpatient

RETURN TO Dr. ____________ IN RADIOLOGY AFTER COMPLETING PART II

NOT A PART OF MEDICAL RECORD

for assessing probabilities. When they returned to their native heaths, they recruited attending physicians from among those who frequently requested them to perform radiological services. They trained the attending physicians in how to estimate probabilities. Under the circumstances the relatively high quality of the probability estimates obtained is surprising and delightful.

The sampling procedure used in this study, like that used in many other studies of medical practice, has one overriding principle; those who participated were those who were willing to participate. No apologies for this procedure are required, since there is no very satisfactory way of preceeding otherwise. Nevertheless, such sampling does present possibilities of bias in generalization to a national population either of radiologists or of attending physicians. Consequently, pending the outcome of further detailed analyses now in progress intended to explore the possibility of sample bias, generalizations from these results to such national populations should be done with extreme caution and nontrivial amounts of skepticism.

Various procedures explained in detail in Lusted et al. (in press) were used to spread cases widely over 47 different emergency rooms and about the same number of radiologists, between large and small hospitals, between teaching and non-teaching hospitals, and over a wide variety and number of attending physicians.

As of July, 1976, the data base was distributed over X-ray procedures as is shown in Table 1.

As usual in any kind of statistical study, there are technical problems, and I must discuss one: the truncation effect. Some respondents responded in probabilities and some responded in odds, but either way most of them worked with relatively small numbers of discrete levels of the quantities they were estimating. In the middle range of uncertainty, this hardly matters, but the extreme ends of

Table 1

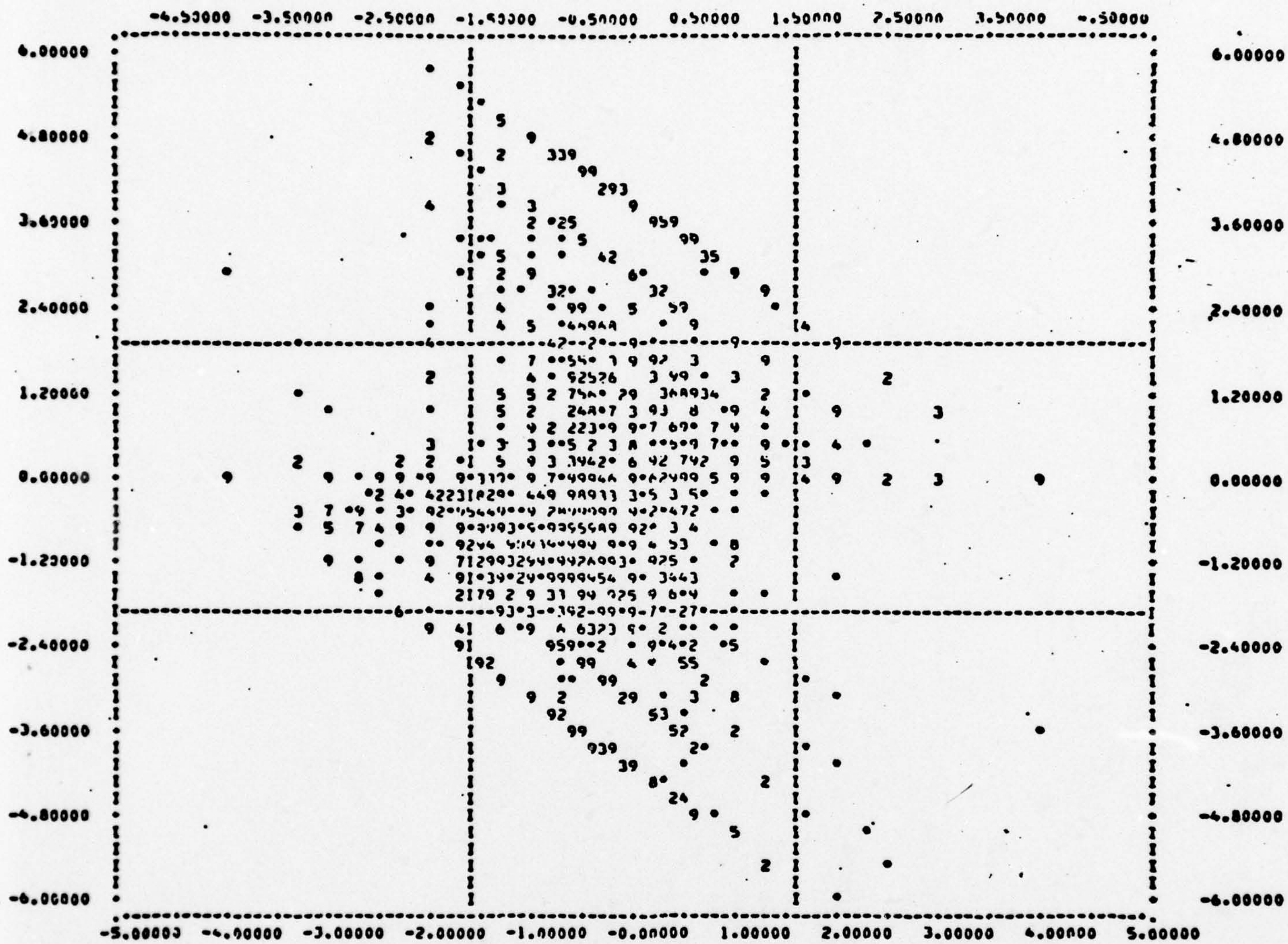
Distribution of Cases Over Procedures

Procedure	Number of Cases
Skull	958
Cervical Spine	862
Chest	2353
Abdomen	839
Intravenous Pyelogram	278
Lumbar Spine	708
Extremities	1878
TOTAL	7876

the scale required particular attention. The problem is more severe for clinicians who reported in probabilities. Many of these, in spite of emphatic attempts to train them otherwise, made estimates of 0 or 1; both of those numbers are uninterpretable in Bayesian arithmetic. ACR adopted an editing convention of calling 0, .0001 and calling 1, .9999. These rounding conventions, combined with the fact that most attending physicians responded in probabilities and used only discrete sets of numbers, produced rather peculiar structures in the analyzed data. Figure 3 presents a scatter plot of log likelihood ratio against log initial odds over all procedures. You can see several parallelogram patterns that correspond to different common truncation limits used by groups of attending physicians, or imposed by the editing convention to avoid estimates of 0 or 1. ACR has, of course, devised methods of analysis that are insensitive to what happens at the extremes of the probability scale. For a more detailed discussion of this technical topic, see Lusted et al. (in press).

Although the study is far from complete, it is possible to base some reasonable convincing conclusions on the data so far. First, the procedure is feasible; that is, such probabilistic assessments can be made in an orderly way and do provide information about the diagnostic thinking of attending physicians. This conclusion follows less from data analysis than from informal contact with the physicians who in fact made the assessments.

A second conclusion is that the impact of X-ray examinations on diagnostic thinking was evident in the vast majority of cases and was substantial in most. Overall, not more than 10% of examinations seemingly had no influence on diagnostic thinking (that is, produced a 0 log likelihood ratio). A more detailed and refined analysis of the data suggests that the actual percentage of 0-information X-rays may be less than 5%.



LOG INITIAL ODDS

Figure 3. Log Likelihood ratio as a function of log initial odds. Physician responses in 4005 cases for all seven radiologic diagnostic procedures.

A third conclusion is that at the time x-rays were requested, the requesting physician was normally uncertain about the correctness of this tentative diagnosis. About 4 times in 5, however, the probability of the tentative most important diagnosis was assessed at less than 1/2; over half the time, it was assessed at less than about .15. In other words, the most important diagnosis often had the character of a not-very-likely medical disaster.

A fourth conclusion is that about 3/4 of the examinations produced a lowering of the clinician's initial probabilities for the tentative most important diagnosis. In other words, on the whole, the effect of radiology in the emergency room setting tends to be one of reassurance rather than one of confirming alarm. This conclusion has implications for the relationship between Efficacy-1, diagnostic efficacy, and Efficacy-2, treatment efficacy. Reassurance is clearly just as appropriate from the point of view of Efficacy-1 as would be confirmation of one's worst fears. On the other hand, it seems quite likely that this finding might imply that x-ray procedures that are highly Efficacious-1 may not be especially Efficacious-2. ACR proposes to attack that question in later studies, if current rather tentative ideas about how to measure Efficacy-2 turn out to be in fact workable.

A fifth conclusion is that the major effect of x-rays is to reduce uncertainty. This was no surprise. Even after examination, however, nearly 40% of clinicians assess probabilities for the most important tentative diagnosis at more than .02 but less than .98. This suggests that a substantial fraction of diagnostic decisions in the emergency room setting are based on weight of evidence rather than proof beyond reasonable doubt. Table 2 shows for various x-ray procedures the percentage of cases with log odds that are either less than -1.75 or greater than +1.75. Those numbers correspond to probabilities of .02 and .98 respectively.

An interesting sixth conclusion, at least from the

Table 2

Percentage of Cases with Log Odds

Less than -1.75 or Greater than +1.75

Procedure	Before Radiography	After Radiography	Net Increase
Skull	15.9	69.9	54.0
Cervical Spine	20.8	77.4	56.6
Intravenous Pyelogram	6.1	54.7	48.6
Lubar Spine	15.8	74.2	58.4
Chest	8.4	55.0	46.6
Abdomen	5.9	45.7	39.8
Extremities	8.4	75.8	67.4
All Procedures (7876 cases)	11.0	65.0	54.0

study so far, is that the influence of X-ray examination on diagnostic thinking was broadly similar for interns, resident physicians in training, and practicing physicians. Also other characteristics, such as the distribution of initial probabilities for diagnoses and the use of odds or probabilities in the expression of uncertainty, were similar for the three groups.

Some other conclusions can be reached from the data, particularily having to do with the question of how well attending physicians used the probabilities they estimated to express their uncertainty. Since these are highly technical in character, I will not review them. I will only add that in general, attending physicians tend to overassess the probability of the relatively unlikely medical disasters that were usually taken as most important diagnoses. Exactly the same kind of finding of overassessment of the probability of highly undesirable events has occurred in a number of other contexts in which probability estimators have the opportunity to confuse their judgments of probability with their assessments of the value of the consequence of the event whose probability was being judged (see Kelly and Peterson, 1971).

A final implication of the study may surprise some. One of the questions asked on the initial form was whether or not the X-ray study was being performed for medical-legal reasons. This box was sometimes checked and sometimes not. Though minor differences between the results when it was checked and when it was not did occur, their smallness was quite surprising. In general, X-rays taken for medical-legal reasons are fully as Efficacious-1 as X-rays for which the attending physician does not indicate that he has such reasons in mind.

How does this study bear on public policy? At the moment, it has no direct bearing. It does suggest that the methodology used is in fact useable, and yields significant information about the behavior of the individuals performing socially important and policy-relevant functions. It is conceivable that refinements of the same methods, combined with methods for measuring Efficacy-2 and perhaps even Efficacy-3, might lead to policy-relevant recommendations about the conditions under which it is or is not most advisable to recommend that X-rays be taken. If such a happy result were to occur, the potential for improving the distribution of health care services might be significant.

Beyond that, however, there is a much more general implication of the study. It shows that decision makers, in this case attending physicians, can and will, with a little training and encouragement, make probability assessments concerning the issues with respect to which they are making decisions. Since uncertainty enters into every decision and probability is the appropriate metric by means of which to quantify uncertainties, this means that the hope of assessing the probabilities that enter into decisions affecting public policy may not be a vain one.

This assertion need not rest solely on this particular study. Many other decision makers besides physicians must deal with uncertainty and are in process of finding the explicit use of probabilities a helpful tool for doing so. Probabilistic weather forecasting is coming to be more and more widely performed. (See for example Murphy and Winkler, 1974). Even more interesting, at least to me, is the growth in use of explicit probabilities among public officials responsible for providing informational input to decision makers concerned with vast issues of global public policy--a growth stimulated mainly by ARPA-sponsored research and application work. For public discussions of relevent technolgy see Edwards, Phillips, Hays, and Goodman (1968), Kelly and Peterson (1971), Barclay and Randall (1975).

In sum, then, Director Dubious, eager to come to terms not only with his own uncertainties but with the uncertainties of those who advise or attempt to influence him, has available to him a quite elaborate technology, based on explicit assessment of probabilities. That technology is already in use, and its generality and simplicity invites optimists like me to suppose that use may extend and spread into other contexts. Perhaps Director Dubious can be helped to become at least somewhat less dubious about uncertainties.

Multiattribute Utility Measurement as a Tool for the Explication and Aggregation of Social Values

As I read Mr. Coates's discussion of the latent, dark, uncongenial, and even unspeakable nature of private motives, I was quite unclear whether he considered this to be desirable, deplorable, or simply a fact of life. But since I don't believe Mr. Coate's premise about the unattractive character of private motives, whether that premise is desirable or deplorable seems to be beside the point. Most motives, public or private, are mundane, ordinary, and reasonably well organized toward the problem at hand. My own motives in deciding what to include in this paper, for example, are to present two intellectual tools that I think may be useful to public decision makers in as effective a light as I can manage, and in the process to be entertaining and perhaps to get a gentle argument going with Mr. Coates. Behind those surface motives, I may well have better-concealed motives to the effect that if the technologies that I am advocating are in fact perceived as useful, I may gain in various ways. None of these motives seem too latent, dark, or uncongenial; and I can guarantee that they are not unspeakable, since I just spoke (or at any rate wrote) about them. Many, perhaps most, of the motives that affect ordinary executives in their working lives have essentially this character.

Mr. Coates made eloquent reference in his paper to the two real problems about motives. One is that different people, and especially different pressure groups, have different motives, whereas the decision maker must make a decision that is responsive both to wishes of those whom he serves and to the technological facts of his problem. The other is that any single person's motives, whether private or public and whether latent or explicit, are virtually always in conflict. And, of course, every public policy decision requires value trade-offs. In order to do better with respect to some dimensions of value, we must do worse with respect to others. But what are the appropriate exchange rates?

A new technology of value trade-offs has been developing very rapidly over the course of the last nine years. It is called multiattribute utility measurement, and it is particularly prominent in the writings of Howard Raiffa, Ralph Keeney, Ron Howard, and myself. Relevant references include Raiffa (1969), Keeney and Raiffa (1976), Howard (1973), and Edwards (1971). ARPA has extensively supported research and applications concerned with this technology and other DoD agencies have applied the technology also. See for example Edwards and von Winterfeldt (1973); Edwards and Gardiner (1975); Edwards and O'Connor (1976); Edwards (1977); Chinnis, Kelly, Minckler, and O'Connor (1976); and O'Connor, Reese and Allen (1976).

The essential idea of multiattribute utility measurement is that every significant value can in effect be partitioned into a set of subvalues on each of a number of dimensions. Technological devices exist for ascertaining what those dimensions are, for locating each one of the actions, objects, or whatever is being evaluated on each of these dimensions, for judging how imprtant each dimension is to the aggregate value of the thing being evaluated, and then for performing the aggregation. Details of this technology vary substantially from one of its advocates to another, but the description as

I have just given it would probably be agreed to by all.

As in the case of probabilities, I intend to review an application that has potential public policy relevance rather than an application in being. There are in fact several applications already in being, and they have been described in open literature. However they are quite complicated. Two examples are: Chinnis, Kelly, Minckler, and O'Connor (1976); and O'Connor, Reese, and Allen (1976). See also Edwards, Guttentag, and Snapper (1975), and Keeney and Raiffa (1976). The particular application that I intend to discuss is to the selection of nuclear waste disposal sites. The work was performed in collaboration with Dr. Harry J. Otway, who is Director of the Research Project on Technological Risk Assessment, sponsored by the International Atomic Energy Authority and the International Institute for Applied Systems Analysis. For a more complete report of this study, see Otway and Edwards (1977).

Otway's project has two main goals. One is to measure the attitudes of various publics toward the risks associated with various modern technologies in general, and with nuclear power production technology in particular. The other is to explore methods by means of which the technological decision makers who must manage nuclear power activities can be aided in taking public attitudes into account in their decision. This particular study was addressed to the latter question. The study was conducted during the course of an international meeting of high level technologists concerned with the problem of nuclear waste disposal. The ten participants included representatives from eight countries with advanced nuclear energy programs. Since the conference was in part about problems of risk assessment and risk management in nuclear waste disposal, they were very much concerned with the problem and very cooperative. Otway planned the study, enlisted the cooperation of the respondents, and collected the data.

The first task, of course, was to find what dimensions of value were relevant to the problem of selecting waste disposal sites. Since Otway's goal was to demonstrate how to take social attitudes toward those sites into account in the decision process, obviously, social attitudes had to be one such value dimension, and indeed it was the first one listed.

Elicitation of value dimensions was done by simply asking all the respondents, together in a room, to identify what issues seemed to them important in making such decisions. Table 3 shows value dimensions and measures for six sites. After Otway had suggested social attitudes as the first such dimension, there was some question about how such attitudes should be scaled, and it was agreed that for the purpose of this demonstration a simple 0 to 100 scale would be appropriate with 100 as a highly favorable attitude and 0 as a highly unfavorable one.

The next dimension, proposed by one of the participants, was remoteness of the waste disposal site from a population center, measured in kilometers. 160 kilometers was considered as having a value of 100 and 0 kilometers was considered as having a value of 0. The third dimension was the geospheric path length in kilometers. Roughly, that is the distance a radioactive particle must travel, typically through the ground, to reach the nearest point used by people. Again, 160 kilometers scores 100 and 0 kilometers scores 0. The fourth dimension was proximity of the waste disposal site to natural resources such as mines. 160 kilometers scores 100 and 0 kilometers scores 0. The fifth dimension was geological disturbance probability--the probability of one or more significant-sized-earthquakes in a year. 10^{-6} (one chance in a million) scores 100 and 1 scores 0. The sixth dimension was the relative migration rate of the critical nuclide, in the geological formation, allowing for absorption and desorption, compared with the rate of movement of ground water (assumed constant at 0.3 m/day). Since this dimension is a ratio, it has no units; 10^{-5} was scored as 100 and 1 was scored as 0.

Table 3

Descriptions of Six Hypothetical Nuclear Waste Disposal Sites

	Value Dimension, Range, and Scaling	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
D1.	Public attitude, 0 = extremely negative; 100 = extemely positive	40	20	10	40	60	70
D2.	Remoteness from Population center, km (90 km = 0; ;60 km = 100)	40	12	12	12	40	120
D3.	Geosperic path length, km (0 km = 0; 160 km = 100)	40	12	12	4	4	40
D4.	Promiximity to natural resources, km (0 km = 0; 160 km = 100)	50	150	150	50	15	15
D5.	Geologic disturbance probaility per year (1 = 0; 10 = 100; linear in exponent)	10^{-4}	10^{-5}	10^{-4}	10^{-6}	10^{-5}	10^{-6}
D6.	Relative migration rate of critical nuclide (1 = 0; 10 = 100; linear in exponent)	10^{-3}	10^{-3}	10^{-2}	10^{-1}	10^{-2}	10^{-1}
D7.	Transportation distance, km (1600 km = 0; 0 km = 100)	1500	500	500	1500	150	150

Table 4

Rescaled single-dimension utilities and aggregate utilities

at six nuclear waste disposal sites

Dimensions	Sites					
	S1	S2	S3	S4	S5	S6
Public attitude	50	16.7	0	50	83.3	100
Remoteness from popluation center	25.9	0	0	100	25.9	100
Geosperic path length	100	22.2	22.2	0	0	100
Proximity to natural resources	25.9	100	100	25.9	0	0
Geological disturbance probability per year	0	50	0	100	50	100
Relative migration rate of critical nuclide	100	100	50	0	50	0
Transportation distance	0	74.1	74.1	0	100	100
Aggregate utility ($\sum_i W_i u_{ij}$)	45.6	57.3	40.4	38.2	41.0	57.9

The seventh dimension, elicited from the respondents only after a great deal of struggle and effort, was transportation distance between the nuclear plan and the waste disposal site. Zero kilometers scores 100 and 1,600 kilometers scores 0.

Note that all dimensions are transformed onto the 0 to 100 scale in such a fashion that higher scores are preferable to lower ones. The scaling of the dimensions was chosen in such a way that the respondents seemed likely to be willing to treat the single dimension utilities as linear with the physical measures involved--and indeed they were. In the case of dimension 5 and dimension 6 this linearity is, of course, with the exponent rather than with the number itself.

In retrospect, several features of the scaling of the dimensions were questionable. The most obvious is the use of 1 as the highest probability of an earthquake in a year. No one would seriously propose a nuclear waste disposal site with so high a probability of an earthquake; a lower probability should have been used as the upper bound.

It is important to emphasize that all sites were assumed to have the same biological characteristics, and that use of any of them was assumed to fall within appropriate budget constraints.

The value model to be used in this particular exercise was a simple weighted average model. Such value models are quite common, and have been exposed to a great deal of criticism by decision analysts (e.g. Keeney and Raiffa, 1976) who complain, quite correctly, that they do not capture subtleties in the value structure that people may bring to a problem. Those like myself, who like to use simple structures and who feel that the simplicity of eliciting numbers built around those structures is more important than getting the model structure just right at the cost of enormously enhanced complexity of elicitation technique, are happy that a number of approximation theorems show that value structures elicited in the way will, under conditions such as prevailed in this experiment, often be very close approximations to much more elaborate and sophisticated value structures that would have required very much more difficult, complicated and socially unacceptable judgments. See Yntema and Torgerson (1961); Dawes and Corrigan (1974); Wainer (1976); and von Winterfeldt and Edwards (1973 (a), 1973 (b)).

In order to perform a simple evaluation of this kind, the next necessary step is to obtain the weights that are to be associated with the various dimensions. The procedure for doing this that I have developed in the course of past ARPA research (Edwards, 1972) is to ask each respondent, working separately, first to rank the dimensions in order of importance, from most to least important. Then he arbitrarily assigns an importance weight of 10 to the least important dimension, and then moves up through the dimensions making ratio judgments about the relative importances of each of the more important dimensions compared with the least important dimension. Since he can also make ratio judgments of the various dimensions compared with one another, he can obtain a great many internal consistency checks to make sure that he is in fact not unduly succumbing to whole number tendencies or any of the other vices to which this kind of judgmental procedure is subject. This was done for each respondent.

Finally, in order to see whether the apparatus that thus had been developed for assessing the attractiveness of waste disposal sites was appealing to the respondents, it was necessary actually to consider some waste disposal sites. So far, the entire process had been carried out without reference to any specific site. However, a number of sites that have been proposed as possible ones for nuclear waste disposal were used as the basis for judgment on the seven relevant dimensions, and the result is shown in Table 3. The ranges of the various dimensions that were actually encountered in the sites were much smaller than the ranges that had been anticipated as possible; this fact has important methodological consequences which I will discuss in a moment.

So far as the respondents were concerned, the final procedure was to ask them to make holistic evaluations, which means ratings on a 0 to 100 scale, of each site, for comparision with the multiattribute utility evaluations.

Otway asked each respondent to judge the importance weights of the seven value dimensions twice and consequently test-retest reliabilities of these judgments could be calculated. Correlations between first and second judgments were very high; the mean was .93. For convenience, all subsequent calculations used the second set of weights. The interrespondent agreement about importance weights was, as you would expect, much lower. Correlations among second judgment weights between pairs of respondents range from +.97 to -.27, with a mean of +.39. Actually, this is a somewhat higher level of inter-judge agreement than has been found in some other applications of this particular technique (e.g. the OCD example in Edwards, Guttentag, and Snapper, 1975). I have argued elsewhere on the basis of ARPA-sponsored research and other data (Edwards (1971); Edwards, Guttentag, and Snapper (1975)) that individual differences in values should show up primarily in assessments of the importance of value dimensions. Single-dimension utilities are often technical judgments rather than value judgmnts.

Obviously, the question that would be of primary interest to Mr. Coates, and also to me, is: How do we go about reducing, removing or otherwise dealing with these individual differences in values?

At this point, unfortunately, time pressure problems arose. The best way to do it would be to normalize the importance weights for each individual separately, to average them, to calculate the ratios of importance weights specified by the averages, and then to feed those ratios back to the judges, sitting as a group, and ask them to debate them until they reach some form of agreement about a final set of such judgments that they were willing to allow to be used in a decision process. The judgments were indeed normalized and averaged, but Otway could not feed back and reconcile differences. In a different context, I have tried this process of feeding back and reconciling differences, with quite good results. And I would anticipate that some procedure of that sort would be the essential ingredient in any large-scale application of this technology to decisions over which there are major social conflicts. In the contexts which the technology has so far been applied, however, the issues involved have been so profoundly technological that such a procedure has not generally been used. Instead, the experts on each of the kinds of numbers were asked to reach consensus about the numbers within the field of their expertise, and were usually able to do so quite well. Perhaps this technology is more easily applicable to fields in which this kind of technological resolution of conflict is appropriate than it is to contexts involving broader kinds of social conflicts.

Now consider the range problem that I mentioned earlier. Consider, for example, dimension 3, geospheric path length. Its actual range covers only 22.5% of the range that originally had been assigned to it. This can easily happen in situations, such as this one, in which the evaluation scheme is developed before the entities to be evaluated are known. Yet exactly that must often be done.

The reason why this presents a problem is that the range of utility values of a value dimension is in a sense a kind of importance weight. A dimension whose utility values range from 0 to 50 is effectively only half as important in controlling evaluation as one having the same weight whose utility values range from 0 to 100.

This problem can be solved only by judgmental methods.

However, some mathematical techniques exist that help to put it into perspective. It is possible to transform both the single-dimension utility values and the importance weights in such a fashion as to preserve unchanged the preference ordering over the options and the utility spacing between options, while putting all of the single-dimension utility functions on a scale whose minimum in fact falls at 0 and whose maximum in fact falls at 100. Table 4 shows the result of doing so. Inspection of that table will show that no one could possibly pick site 3. In technical jargon, site 2 dominates site 3; that is, site 2 is at least as good as site 3 on every dimension, and definitely better on at least one. No other site is dominated. Also note that site 6, although evaluated as best by the weighted utility criterion, does not dominate site 3; site 3 is better than site 6 on the dimensions of proximity to natural resources and transportation distance.

The transformations which I have discussed permit exploration of the extent to which the scaling of the single dimension utility functions influences the ultimate outcome. I won't go into the details, but in this particular instance, which is rather extreme in deviations of the actual from the anticipated ranges, the effect on preference orderings was extremely modest. In other words, this procedure is rather robust to errors of anticipation of that sort.

Finally, consider the relation between the holistic ratings for the other sites by the respondents and the multiattribute utility ratings. The mean correlation in holistic ratings between pairs of respondents is +.20, and the range is from +.97 to -.55. Note that the respondents are even less in agreement about holistic ratings than they were about importance weights. That too is a common finding in applications of this method. The correlation between mean holistic ratings and multiattribute utility ratings is +.58. Both procedures consider site 6 to be best and site 3 to be worst. This correlation between multiattribute utilities and holistic ratings is somewhat high compared with most other such correlations in the multiattribute utility literature, although it still shows that the two procedures do lead to different results. That on the whole is gratifying. After all, there would be no point in procedures like multiattribute utility measurement if direct numerical assements produced exactly the same results. Except for various technical details having to do with intercorrelations among dimensions, both in value and in physical characteristics, and with the effect of these on scaling procedures, that's the end of the story of this particular study, except for one important addition. Harry Otway informs me that the respondents thoroughly enjoyed the study, found the importance weights that they had judged extremely enlightening, and requested him to be prepared to repeat the study at their next meeting, with a considerably more realistic setting and paying considerably more attention to the details of how the study is done.

Much more sophisticated and complicated versions of exactly the same technology have been used and are now being usedunder ARPA and other DoD sponsorship to make major socially important decisions. Several have been published in unclassified sources. For example, one (Chinnis, et al., 1976) has to do with the selection of the winning bidder from among a number of bids in a very large-scale procurement of an important and expensive item of military hardware. The additional complexities of the method were concerned primarily with the much larger number of dimensions that were taken into account, the use of a hierarchical value model rather than the simple value model I have presented here, and the introduction of scenarios and scenario probabilities as a tool for the assessment of values. While these technological details are all of fundamental importance to real applications, nothing in them changes the basic idea I have present in this rather simple-minded exposition.

Nor are all the examples military. In one published application (Edwards, Guttentag, and Snapper, 1975) a technique of essentially this character was used to help a major agency within the Department of Health, Education, and Welfare to make decisions about the allocation of its research budget for a year. In another application, about to appear in the ARPA technical report, the same kind of technology is being used in planning the rate at which a government agency should encourage a boom town to boom. Still another application now in progress is to the National Program for Decriminalization of Status Offenders. A great deal of data has been collected by Professor Solomon Kobrin and his collaborators at the Social Science Research Institute of USC on the impact of this program both on the juveniles with whom it deals and on the criminal justice and related agencies who must deal with these juveniles. We are now collecting multiattribute utility measurements from a number of experts on juvenile deliquency, crime, the juvenile justice system, and the like, and expect to use these judgments in the process of assessing what the overall effects of this major national program in fact have been, and whether those effects are good or bad, and how good or how bad.

Conclusion

This paper, after some initial questioning of the assertion that major issues of public policy are inaccessible to technological tools, has attempted to illustrate the nature of two technological tools, and to suggest how they can be and are being used in the course of making major social policy decisions. Obviously, I would not want to claim that these tools are optimal, that they are fully developed, or that they should be used for all such decisions. Their applicability is quite limited, as I have attempted to suggest in the course of sketching their nature. Within that area of applicability, however, I believe that they can help those charged with responsibility for social policy in dealing with the two key problems that Mr. Coates identified: uncertainty, and difficulties in assessing and reconciling values.

As Mr. Coates correctly pointed out, no technological tool is likely to be of very great use to Director Devious. His conception of his function, and his goal structure, makes him essentially uninfluenceable by the technology of decision making. Indeed, only the part of that technology that has to do with budgeting and the assessment of costs is likely to get very much of his attention.

On the other hand, as I suggested at the beginning of this paper, Director Dubicus is less impervious, mostly because he is less convinced that social policy making must continue to be done in the way in which it always has been done. I conceive of Director Dubious as a skeptical but open-minded man, interested in technological innovation and willing to explore the possibilty that a particular technological innovation may have something useful to offer him. I have suggested two possible candidate technologies for his attention.

References

Barclay, S. and Randall, L.S. Interactive decision analysis aids for intelligence analysts. Technical Report DT/TR 75-4. McLean, Va.: Decisions and Designs, Inc., December, 1975.

Bell, R.S., and Loop J.W. The utility and futility of radiographic skull examination for trauma. New England Journal of Medicine, 1971, 284, 236-239.

Chinnis, J.O., Kelly, C.W., III, Minckler, R.D., and O'Connor M.F. Single channel ground and airborne radio system (SINCGARS) evaluation model. Technical Report DT/TR 75-2. McLean, Va.,: Decisions and Designs, Inc., August, 1976.

Coates, J.F. What is a public policy Issue? Paper to appear in AAAS Symposium Volume on Judgment and choice in Public Policy Decisions. (In Press).

Edwards, W. Social Utilities. In Decision and Risk Analysis: Powerful New tools for management, Proceedings of the Sixth Triennial Symposium, June 1971, Hoboken: The Engineering Economist, 1972, 119-129.

Edwards, W. How to use Multiattribute Utility Measurement for Social Decision Making. In IEEE Transactions on Systems, Man, and Cybernetics. May 1977, 7, 5, 326-340.

Edwards, W., Lindman, H., and Phillips, L.D. Emerging technologies for making decisions. In New Directions in Psychology II. New York: Holt, Rinehart and Winston,1965.

Edwards, W., Phillips, L.D., Hays, W.L., and Goodman, B.C. Probabilistic information processing systems: Design and evaluation. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 248-265.

Gardiner, P.C. and Edwards,W. Public Values: Multiattribute Utility Measurement for Social Decision Making. Kaplan, M.F., and Schwartz, S. (Eds.) In Human Judgment and Decision Processes, New York: Academic Press, 1975.

Howard, R.A. Decision analysis in systems engineering. In Miles, R.F., Jr., (Ed.), Systems concepts. New York: Wiley, 1973.

Keeney, R.L., and Raiffa, H. Decisions with multiple objectives: Preferences and value tradeoffs. New York: Wiley, 1976.

Kelly, C.W. III, and Peterson, C.R. Probability estimates and probabilistic procedures in current-intelligence analysis. IBM Rep. 71-5047. Gaithersburg, Md.: International Business Machines, 1971.

Lusted, L.D., Bell, R.S., Edwards, W., Roberts, H.V., and Wallace, D.L. Evaluating the efficacy of radiologic procedures by Bayesian Methods: A progress report. In Snapper, K. (Ed.), Models and metrics for decision Makers. Washington, D.C.; Information Resources Press, in press.

Murphy, A.H., and Winkler, R.L. Probability forecasts: A survey of national weather service forecasters. Bulletin of the American Meteorological Society, 1974, 55, 1449-1453.

O'Connor, M.F., Reese, T.R., and Allen, J.J. A multiattribute utility approach for evaluating alternative Naval aviation plans. Technical Report DT/TR 76-16. McLean, Va.: Decisions and Designs, Inc., September, 1976.

O'Connor, M.F., and Edwards, W. On Using Scenarios in the Evaluation of Complex Alternatives. Decision and Designs, Inc. Technical Report 76-17. Sponsored by Defense Advanced Research Projects Agency. ARPA Order No. 3053.

Otway, H.J., and Edwards, W. Application of a simple multiattribute rating technique to evaluation of nuclear waste disposal sites: A demonstration, Vienna, Austria: International Atomic Energy Authority, Research Memorandum, RM-77-31, June, 1977.

Raiffa, H. Preferences for multiattributed alternatives. RM-5868-DOT/RC. Santa Monica, Ca,: The Rand Corporation, 1969.

Thornbury, J.R., Fryback, D.G., and Edwards, W. Likelihood ratios as a measure of the diagnostic usefulness of the excretory urogram information. Radiology, 1975, 114 561-565.

von Winterfeldt, D. and Edwards, W. Costs and payoffs in perceptual research. University of Michigan, Engineering Psychology Laboratory Report 011313-1-T, October, 1973. (a)

von Winterfeldt, D., and Edwards, W. Flat maxima in linear optimization models. University of Michigan, Engineering Psychology Laboratory Report 011313-4-T, November, 1973. (b) von Winterfeldt, D. and Edwards, W. Evaluation of Complex Stimuli using multiattribute utility procedures. University of Michigan, Engineering Psychology Laboratory Report 011313-2-T, October, 1973.

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Yntema, D.B., and Torgerson, W.S. Man-computer cooperation in decisions requiring common sense. IRE Transactions on Human Factors in Electronics, 1961, HFE-2,20-26.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D.C. 20375	6 copies
Office of Naval Research Attention: Code 102IP 800 North Quincy Street Arlington, Virginia 22217	6 copies
Decisions and Designs, Incorporated 8400 Westpark Drive, Suite 600 McLean, Virginia 22101	20 copies

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Naval Personnel Research and Development
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N215)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 540
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Alan H. Curry
Operations and Management Science Divisior
U.S. Army Institute for Research in Manage
ment Information and Computer Science
730 Peachtree St., N.E. (Suite 900)
Atlanta, Georgia 30308

Revised December 1977

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense Research and Engineering (Test and Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Assistant Director (Environmental and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: COL Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
APO San Francisco 96610

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations (OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
Research Laboratory
U.S. Army Institute for the Behavioral
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Technical Director, U.S. Army Concepts
Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Commanding General
Headquarters, DARCOM
Attention: DRCED - Richard Murray
5001 Eisenhower Avenue
Alexandria, VA 22333

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office
(Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
and Acquisition Programs
Office of the Deputy Chief of Staff for
Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Director, Advanced Systems Division
(AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 001855-2-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Technology for Director Dubious; Evaluation and Decision in Public Contexts		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 10/76-9/30/77
		6. PERFORMING ORG. REPORT NUMBER 77-2
7. AUTHOR(s) Ward Edwards		8. CONTRACT OR GRANT NUMBER(s) Prime Contract N00014-76-C-004 Subcontract 76-030-0715
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California Los Angeles, CAlifornia 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE November, 1977
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Decisions & Designs, Inc. Suite 100, 7900 Westpark Drive McLean, Virginia 22101 (Under contract from Office of Naval Research)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

efficacy
log likelihood ratios
multiattribute utility measurements
Bayesian inference
prior probabilities
log final odds
log initial odds

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report examines the social, structural and organizational obstacles to the introduction of decision technology in public contexts, and summarizes two studies that suggest ways of overcoming these obstacles. As a means of defining the problem the report caricatures two Federal government policy-makers: Director Devious and Director Dubious. Director Devious wants to keep his freedom of action. Director Dubious, though a skeptic about new technologies, recognizes the problem they address and is willing to give

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) them a try. Two classes of technological tools are proposed to him. One technology, concerned with probability estimation and Bayesian inference, is illustrated by a study conducted by the American College of Radiology, using ARPA-developed technology, of the diagnostic value of x-rays. Attending physicians, minimally trained about probabilities, made pre and post x-ray probability judgments about possible diagnoses in emergency room cases. The log likelihood ratio inferred from these judgments were the measure of diagnosticity. The main conclusins were: 1) minimally trained physicians make very well calibrated probability estimates, 2) very few x-rays are completely undiagnostic, even if taken for medical legal reasons, 3) level of physician training made little difference to performance in probability estimates. The other technology, concerned with measurement of social values, is illustrated by an application of a version of multiattribute utility measurement to selection of nuclear waste disposal sites. Experts on nuclear waste disposal sites evaluated various hypothetical sites by an ARPA-developed procedure. The main findings were that they liked the procedure and wanted to try it further, and that the results were robust under manipulations having to do with incorrect prior expectations concerning the ranges of dimensions of value. Both technologies are offered to Director Dubious and his governmental colleagues, as serious candidates for adaptation and use.